

---

# A Comparison of Surface and Motion User-Defined Gestures for Mobile Augmented Reality

**Ze Dong**

HIT Lab NZ  
University of Canterbury  
New Zealand  
ze.dong@pg.canterbury.ac.nz

**Thammathip****Piumsomboon**

School of Product Design  
University of Canterbury  
New Zealand  
tham.piumsomboon@  
canterbury.ac.nz

**Jingjing Zhang**

HIT Lab NZ  
University of Canterbury  
jingjing.zhang@  
pg.canterbury.ac.nz

**Adrian Clark**

School of Product Design  
University of Canterbury  
New Zealand  
adrian.clark@  
canterbury.ac.nz

**Huidong Bai**

Auckland Bioengineering  
Institute  
University of Auckland  
New Zealand  
Huidong.bai@auckland.ac.nz

**Rob Lindeman**

HIT Lab NZ  
University of Canterbury  
New Zealand  
gogo@hitlabnz.org

**Abstract**

Advancements in Augmented Reality (AR) technologies and processing power of mobile devices have created a surge in the number of mobile AR applications. Nevertheless, many AR applications have adopted surface gestures as the default method for interaction with virtual content. In this paper, we investigate two gesture modalities, surface and motion, for operating mobile AR applications. In order to identify optimal gestures for various interactions, we conducted an elicitation study with 21 participants for 12 tasks, which yielded a total of 504 gestures. We classified and illustrated the two sets of gestures, and compared them in terms of goodness, ease of use, and engagement. The elicitation process yielded two separate sets of user-defined gestures; legacy surface gestures, which were familiar and easy to use by the participants, and motion gestures, which had better engagement. From the interaction patterns of this second set of gestures, we propose a new interaction class called TMR (Touch-Move-Release), which defines for mobile AR.

**Author Keywords**

Augmented Reality; mobile devices; gestures; surface gestures; motion gestures; elicitation study.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the Owner/Author.

CHI'20 Extended Abstracts, April 25–30, 2020, Honolulu, HI, USA  
© 2020 Copyright is held by the owner/author(s).  
ACM ISBN 978-1-4503-6819-3/20/04.  
<https://doi.org/10.1145/3334480.3382883>

Table 1: Selected tasks.

#	Apps	Tasks
1	Angry Birds	Slingshot
2	Pokémon GO	Throw
3	Just a Line	Draw
4	Just a Line	Erase
5	CoolAR	Scale Up
6	CoolAR	Scale Down
7	IKEA Place	Move
8	IKEA Place	Rotate
9	ARia	Open Drawer
10	ARia	Close Drawer
11	CoolAR	Open Door
12	CoolAR	Close Door

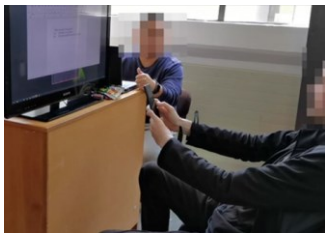


Figure 1: Experimental Setup - a participant is performing a gesture while watching a video displayed on a TV screen.

## CSS Concepts

• **Human-centered computing~Human computer interaction (HCI)** ~Interaction paradigms~Mixed / augmented reality; Interaction techniques~Gestural input.

## Introduction

According to Azuma [1], Augmented Reality (AR) systems must fulfill three requirements, that they 1) combine real and virtual images, 2) are interactive in real-time, and 3) register virtual content in 3D space. Due to their suitability for fulfilling these three requirements and ubiquity, mobile phones have become the primary way for people to experience AR applications encompassing a variety of domains [4, 16]. Mobile AR enabling frameworks, such as Apple's ARKit [7] and Google's ARCore [8], made development of mobile AR applications accessible to more developers than ever, and there are now a wide range of mobile applications which use AR as a core mechanic. For example, IKEA Place [2] allows customers to visualise virtual furniture in their home. QuiverVision [11] was the first to introduce AR coloring books, and SketchAR [13] teaches users how to draw by overlaying virtual drawings over a real canvas.

Nevertheless, most current mobile AR applications adopt existing interaction metaphors based on surface gestures designed for devices with a touch-sensitive screen. Surface gestures may be familiar to the user, however, they have been found to be less engaging when used for AR experiences as the interaction was restricted to a 2D plane instead of making use of the 3D nature of AR [3]. Furthermore, the small screen size of mobile devices means that the interaction is limited and also causes hand occlusion of the display [6].

Past research has proposed methods to enrich mobile AR experiences, for example by combining smartwatches and mobile phone for interaction [9; 14], or by integrating additional sensors with the handset [15]. However, these methods require additional equipment to achieve a better experience, vastly restricting the number of users who can access these experiences. For this reason, our interest is in utilising existing mobile device sensors to create new and engaging interaction methods more suitable for mobile AR experiences. Previous research has shown that complex tasks can be achieved with the common on-device sensors [18], utilising users' mental model [19], and previous elicitation studies such as Ruiz [12] and Piumsomboon [10] have elicited gestures that are performed motion with an arm movement, which demonstrates that some degree of physical exertion might improve the overall experience in mobile AR.

In this paper, we present an elicitation study using the method from Wobbrock et al. [17] with 21 participants for 12 tasks in mobile AR applications. The participants were asked to design two gestures for each task, a surface gesture and a motion gesture, for each task. They were also asked to rate their gestures in terms of goodness, ease of use, and engagement. This yielded a total of 504 user-defined gestures, and after classification and ranking, we selected a final set of 25 gestures comprising of 13 surfaces and 12 motion gestures. From the study, we observed a design pattern for motion gestures, and propose a technique called TMR (Touch-Move-Release) for designing engaging gestures for mobile AR interaction.

## Eliciting Surface and Motion Gestures

This elicitation study followed a similar process to Wobbrock et al. [17]. To come up with appropriate



Figure 3: The number of types of gestures elicited for each task, surface gestures in blue and motion gesture in red.

tasks for the elicitation study, we surveyed 16 mobile AR applications on both Google Play and the Apple App Store. In the end, we selected twelve tasks from six applications (see Table 1), based on their high level of commonality across applications, together with some tasks that also appeared in past research [5, 10, 12, 17]. To elicit the gestures, we prepared a set of videos for the twelve tasks by recording the screen during the interaction from the six chosen applications. Each task contained two videos: 1) a video mimicking surface interaction to manipulate the virtual object with minimal movement of the mobile device, 2) a video mimicking motion interaction with a movement of the mobile device when interacting with the virtual object. After designing a gesture for each video, the participants were asked to rate their gestures on a 7-point Likert scale in terms of *Goodness* - how suitable was the gesture for the task, *Ease of Use* - how easy was it to perform, and *Engagement* - how engaging it was to use. Each task took approximately 4 minutes and the study lasted approximately an hour.

#### Participants

Twenty-one participants (11 males and 10 females) were recruited, aged 18 years to 59 years old, with an average age of 29 (SD=10.7) years. They were all right-handed. All of the participants owned a touch-screen mobile device, however, 8 of them had no prior experience with mobile AR and the remainder had some experience but none were frequent users.

#### Setup

The participants were given a mobile phone, a Samsung Galaxy S9, to hold. To overcome the limited screen size and any issues of occlusion during the gesture while using the phones display, we had our participants seated in front of a 32" television screen to

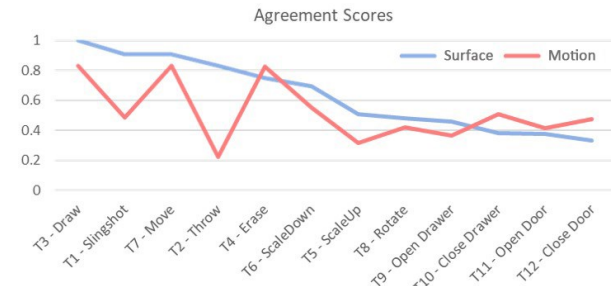


Figure 2: Agreement scores for each task - surface gestures in descending order (blue line) and motion gestures (red line).

watch the videos instead. The participants were asked to follow a think-aloud protocol and their gestures were recorded with a camera placed over their right shoulder.

#### Hypothesis

For surface gestures, which are well established, universal, and highly familiar to regular users of mobile phones and tablets, we expect to elicit a legacy set of common surface gestures being used in mobile AR applications. For the comparison of the ratings for the two sets of gestures, our hypotheses are **H1)** the surface gestures would be rated higher in terms of *Goodness* and *Ease of Use* but **H2)** participants would find motion gestures to be more *engaging*.

#### Result

The study yielded a total of 504 gestures from the 24 videos. Figure 3 shows the number of different types of gestures being elicited. Based on the most extensive set of identical gestures for each task, we constructed the user-defined gestures for the surface and motion interaction as shown in Figure 4, each set contained 13 and 12 gestures, respectively.

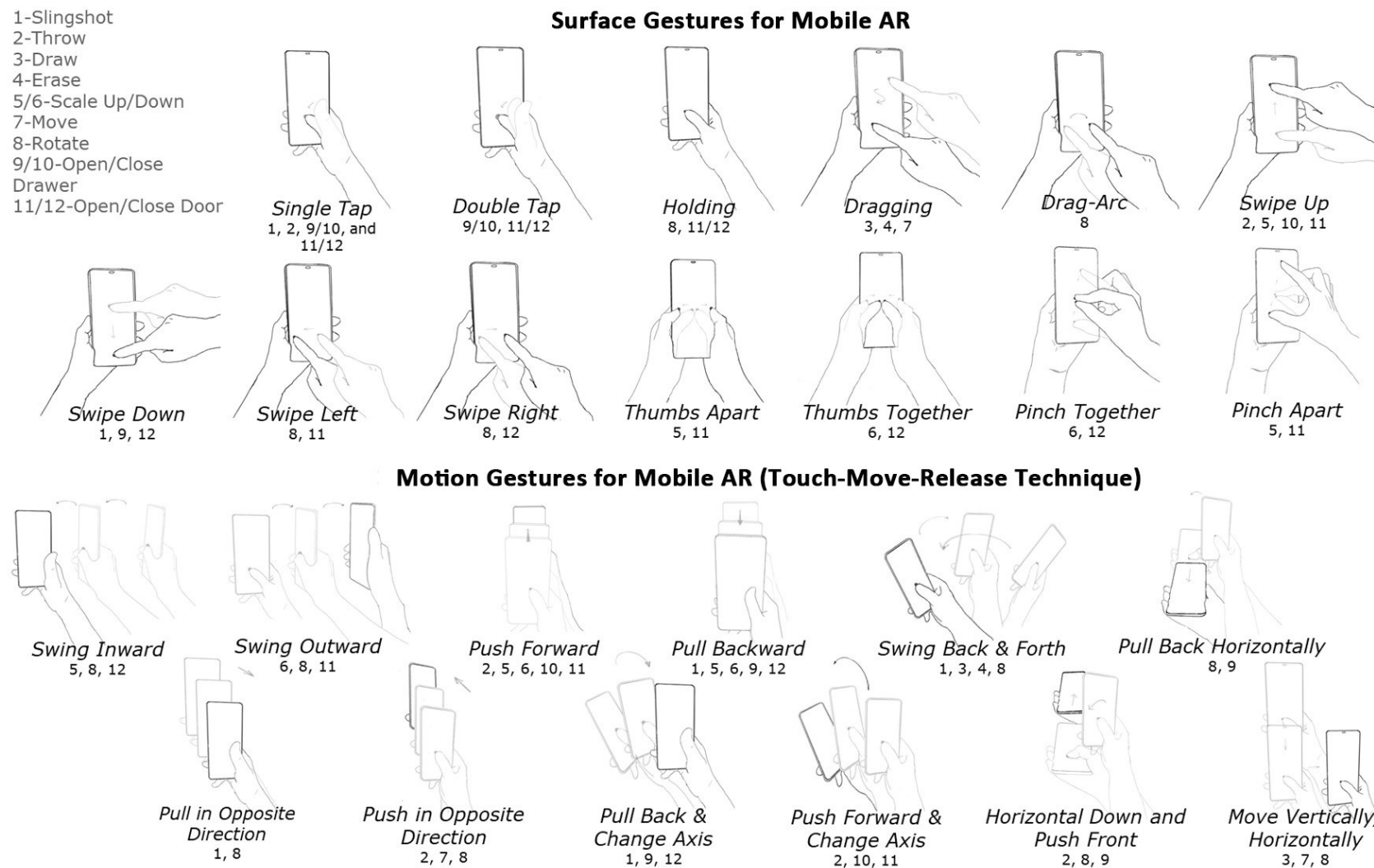


Figure 4: Two sets of user-defined gestures, surface gestures for mobile AR (top), and motion gestures for mobile AR (bottom). Motion gestures demonstrate the concept of Touch-Move-Release (TMR) interaction technique. TMR utilizes device's touchscreen to initiate and terminate interaction and device's AR tracking for arm movements to engage the user with more physical activity in 3D space for better AR experience.

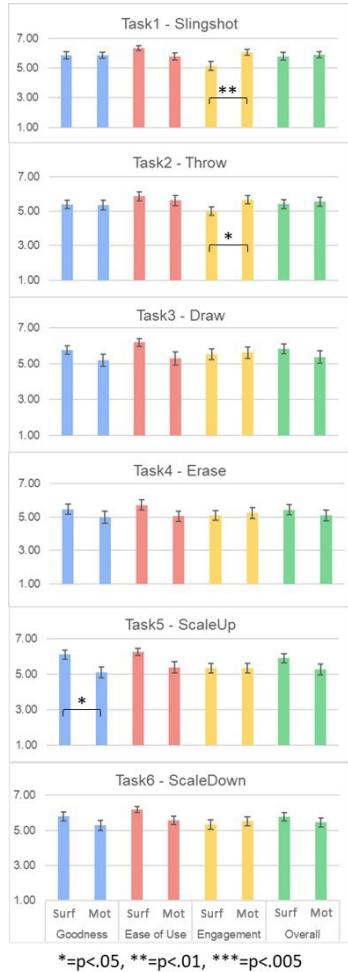


Figure 5: User ratings in terms of *Goodness*, *Ease of Use*, *Engagement*, and *Overall* ratings for task 1 to 6.

### Level of Agreement

The agreement score for each task was calculated using equation 1 for both sets based on [17].

$$A = \sum_{P_s} \left( \frac{|P_s|}{|P_t|} \right)^2 \quad - \text{Equation 1}$$

$P_t$  represents the number of gestures designed for each task, and  $P_s$  is the number of similar gestures in the same task. The results are plotted and illustrated in Figure 2. By comparing the two sets of gestures the difference between the agreement scores was notable for task 1 (0.91-Surface, 0.48-Motion) and task 2 (0.83 vs 0.22).

### User-defined Gesture Characteristics

As shown in Figure 4, multiple gestures could be used to perform multiple tasks. For example, in the surface gesture set, task 9 (*Open Drawer*), 10 (*Close Drawer*), 11 (*Open Door*), and 12 (*Close Door*) shared the “double-tap” gesture. Moreover, for the surface gesture set, “swiping” and “holding” gestures were common occurrences in the interaction design pattern. On the contrary, motion gestures had a generally lower agreement as shown in Figure 2. This was expected as an additional dimension and greater interaction possibilities gave participants a larger design space. Nevertheless, we observed common patterns in the characteristics of the elicited motion gestures. Firstly, the trajectory of the gestures, i.e. device movement’s direction, varied but was generally aligned with the desired movement of the virtual object. Secondly, participants could utilize the touch-sensitive screen to initiate and terminate their action. This observation led us to propose the Touch-Move-Release (TMR) technique, which involves three steps of action

corresponding to the functions of initiating, calibrating, and terminating an interaction.

### Comparisons of Ratings

To validate our hypotheses, we analysed the three rating scores between the surface and motion gestures in terms of *Goodness*, *Ease of Use*, and *Engagement*. The Friedman test followed by a post-hoc pairwise comparisons, Wilcoxon signed-rank tests with Bonferroni correction (with p-value adjusted) were used to compare the two sets of ratings. Figure 5 and 6 illustrate the plots. We use  $\bar{x}_s$  to indicate the mean rating for the surface gesture and  $\bar{x}_a$  for motion, and the 2<sup>nd</sup> number in the bracket after the mean represents SD.

For *Goodness* scores, we found significant differences in task 5-Scale Up ( $V=123.5$ ,  $p=0.03$ ,  $\bar{x}_s=6.1(1.2)$ ,  $\bar{x}_a=5.1(1.4)$ ), 7-Move ( $V=6$ ,  $p=0.03$ ,  $\bar{x}_s=5.8(1.2)$ ,  $\bar{x}_a=6.4(0.6)$ ), 9-Open Drawer ( $V=91$ ,  $p=0.02$ ,  $\bar{x}_s=6.1(1.0)$ ,  $\bar{x}_a=4.7(1.8)$ ) and 10-Close Drawer ( $V=75.5$ ,  $p=0.04$ ,  $\bar{x}_s=6.1(0.9)$ ,  $\bar{x}_a=5.1(1.5)$ ).

For *Ease of Use* scores, significant differences were found for task 9-Open Drawer ( $V=91$ ,  $p=0.001$ ,  $\bar{x}_s=6.7(0.5)$ ,  $\bar{x}_a=5.3(1.6)$ ), 10-Close Drawer ( $V=101.5$ ,  $p=0.002$ ,  $\bar{x}_s=6.6(0.5)$ ,  $\bar{x}_a=5.2(1.6)$ ), 11-Open Door ( $V=85$ ,  $p=0.04$ ,  $\bar{x}_s=6.3(0.8)$ ,  $\bar{x}_a=5.7(1.1)$ ), and 12-Close Door ( $V=78$ ,  $p=0.02$ ,  $\bar{x}_s=6.3(0.8)$ ,  $\bar{x}_a=5.4(1.3)$ ).

For *Engagement*, we found significant differences for task 1-Slingshot ( $V=9$ ,  $p=0.006$ ,  $\bar{x}_s=5.1(1.3)$ ,  $\bar{x}_a=6.1(1.0)$ ), 2-Throw ( $V=39.5$ ,  $p=0.04$ ,  $\bar{x}_s=5.0(1.2)$ ,  $\bar{x}_a=5.7(1.1)$ ), 7-Move ( $V=13.5$ ,  $p=0.005$ ,  $\bar{x}_s=5.3(1.1)$ ,  $\bar{x}_a=6.4(0.7)$ ), and 11-Open Door ( $V=33.5$ ,  $p=0.04$ ,  $\bar{x}_s=4.7(1.7)$ ,  $\bar{x}_a=5.8(1.2)$ ). Finally, the *Overall Score*, which combined all the three scores yielded significant differences for task 7-Move

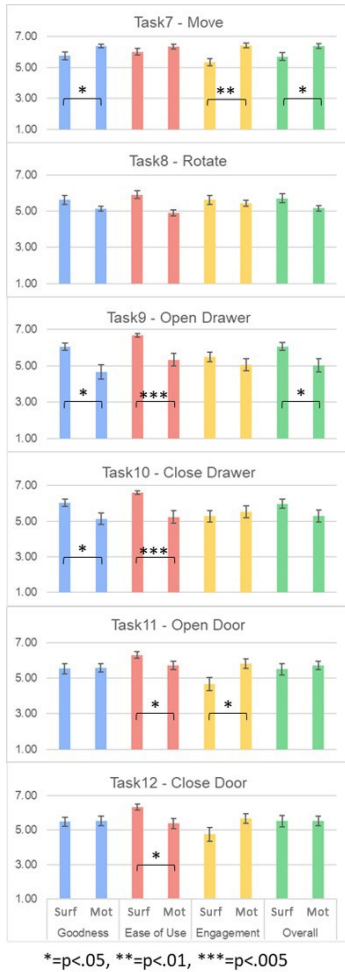


Figure 6: User ratings in terms of *Goodness*, *Ease of Use*, *Engagement*, and *Overall* ratings for task 7 to 12.

( $V=21.5$ ,  $p=0.02$ ,  $\bar{x}_s=5.7(1.1)$ ,  $\bar{x}_a=6.4(0.7)$ ) and 9-Open Drawer ( $V=164.5$ ,  $p=0.03$ ,  $\bar{x}_s=6.1(1.0)$ ,  $\bar{x}_a=5.0(1.7)$ ).

### Discussion

The results support our hypotheses H1 and H2 for some tasks. Surface gestures were rated significantly higher on *Goodness* for tasks 5, 9, and 10, while motion rated higher for task 7. For *Ease of Use*, we found surface gesture were easier to perform for tasks 9 to 12. However, in terms of *Engagement*, motion gestures were rated significantly higher for tasks 1, 2, 7, and 11. The rating outcomes indicated that the participant's perception of their gestures was deeply tied to the nature of the tasks and was more complex than the initial impression of the experimenters. Further examination of the ratings for each task and the resulting gestures yielded new insights.

As we expected, the elicited surface gestures was compromised of a familiar legacy set of gestures. We found the surface gestures were better suited for tasks 5, 9, and 10. It was also found to be easier to perform for tasks 9 to 12. The reason behind this might be due to user familiarity and the lower physical effort it took to perform the task, particularly for binary modality tasks. For example, in the *Open Drawer* task, it did not matter how the drawer was opened, whether it was opened slowly or at half the distance, and therefore, any gesture which executed the action would satisfy the goal.

On the contrary, motion gestures were found to be more engaging for the interactions, such as using a slingshot, throwing, moving, and opening a door. This was expected as the gestures reflected the real-world physical movement involved in performing similar physical tasks in the real world. From the *Overall*

results, surface gestures were rated better for tasks such as *Open Drawer*, but motion gestures were more suitable for tasks which may require finer control such as moving in 3D space. Participant feedback indicated the majority of users preferred motion gestures if they were used in the game context. However, for non-game mobile AR applications, participants preferred using surface gestures which take less effort to perform. From these findings, we propose using TMR gestures to improve the level of engagement in mobile AR applications especially in a gaming context.

### Limitations

The findings led us to summarize the drawbacks of TMR technique as follows: 1) Visual Impairment – the display might not be visible to the user at all time, 2) Cybersickness – fixating on the screen while moving might cause sickness, 3) Fatigue – requiring more physical movement which might not be ideal for experiences of longer duration.

### Conclusion and Future Work

We conducted a study to elicit two sets of gestures, surface and motion, targeting mobile AR experience. We surveyed and selected 12 tasks from 6 mobile AR applications. The study yielded 504 gestures, which we selected a total of 25 gestures in the final user-defined gesture set, including 13 surface and 12 motion gesture. We compared the two sets of gestures in terms of *Goodness*, *Ease of Use*, and *Engagement*. We found that surface gestures elicited were familiar legacy gestures that were easy to use but motion gestures were more engaging. We propose Touch-Move-Release (TMR) technique to improve engagement for mobile AR applications.



### Acknowledgement

This work is supported by National Natural Science Foundation of China (NSFC), Grant No: 61850410532.

### References

- [1] Ronald Azuma. 1997. A survey of augmented reality. *Presence: Teleoperators & Virtual Environments* 6, 4, 355-385.
- [2] IKEA Systems B.V. 2019. IKEA apps - IKEA inter IKEA Systems B.V., ikea.com.
- [3] Huidong Bai. 2016. Mobile Augmented Reality: Free-hand Gesture-based Interaction.
- [4] Mark Billinghurst, Adrian Clark, Gun Lee. 2015. A survey of augmented reality. *Foundations and Trends® in Human-Computer Interaction* 8, 2-3, 73-272.
- [5] Eg Su Goh, Mohd Shahrizal Sunar, Ajune Wanis Ismail. 2019. 3D Object Manipulation Techniques in Handheld Mobile Augmented Reality Interface: A Review. *IEEE Access* 7, 40581-40601.
- [6] Robert Hardy, Enrico Rukzio. 2008. Touch & interact: touch-based interaction of mobile phones with displays. *In Proceedings of the 10th international conference on Human computer interaction with mobile devices and services ACM*, 245-254.
- [7] Apple Inc. 2019. Augmented Reality - Apple Developer, Overview of the ARKit 3.
- [8] Google Inc. 2019. ARCore Overview |Google Developers google Inc., developers.google.com.
- [9] Hyung-il Kim, Woontack Woo. 2016. Smartwatch-assisted robust 6-DOF hand tracker for object manipulation in HMD-based augmented reality. *In 2016 IEEE Symposium on 3D User Interfaces (3DUI)* IEEE, 251-252.
- [10] Thammathip Piumsomboon, Adrian Clark, Mark Billinghurst, Andy Cockburn. 2013. User-defined gestures for augmented reality. *In IFIP Conference on Human-Computer Interaction* Springer, 282-299.
- [11] QuiverVision, 2016. QuiverVision 3D Augmented Reality Coloring Apps QuiverVision, quivervision.com.
- [12] Jaime Ruiz, Yang Li, Edward Lank. 2011. User-defined motion gestures for mobile interaction. *In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems ACM*, 197-206.
- [13] SketchAR. 2019. Start Drawing Easily Using Augmented Reality. SketchAR SketchAR, UAB., sketchar.tech.
- [14] Darko Stanimirovic, Daniel Kurz. 2014. Smartwatch-aided handheld augmented reality. *In 2014 IEEE International Symposium on Mixed and Augmented Reality (ISMAR) IEEE*, 373-374.
- [15] Tomohiro Tanikawa, Hidenori Uzuka, Takuji Narumi, Michitaka Hirose. 2015. Integrated view-input AR interaction for virtual object manipulation using tablets and smartphones. *In Proceedings of the 12th International Conference on Advances in Computer Entertainment Technology ACM*, 7.
- [16] DWF Van Krevelen, Ronald Poelman. 2010. A survey of augmented reality technologies, applications and limitations. *International journal of virtual reality* 9, 2, 1-20.
- [17] Jacob O Wobbrock, Meredith Ringel Morris, Andrew D Wilson. 2009. User-defined gestures for surface computing. *In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems ACM*, 1083-1092.
- [18] Nan Zheng, Kun Bai, Hai Huang, Haining Wang. 2014. You are how you touch: User verification on smartphones via tapping behaviors. *In 2014 IEEE 22nd International Conference on Network Protocols IEEE*, 221-232.

- [19] Kening Zhu, Xiaojuan Ma, Haoyuan Chen, and Miaoyin Liang. "Tripartite effects: exploring users' mental model of mobile gestures under the influence of operation, handheld posture, and interaction space." *International Journal of Human-Computer Interaction* 33, no. 6 (2017): 443-459.